GRITS/STAT 5.0 AND THE YEAR 2000

Doug Kimzey
PC Engineering, Inc.
December 15, 1997

Abstract
This white paper provides a succinct explanation of the use of the **GRITS/STAT 5.0** software and the year 2000 problem. **GRITS SAGE** version 1.1 or higher is currently the only way to enter data with year 2000 sampling dates. **GRITS Database** and **GRITS SAGE** versions 1.0d or lower *do not* allow users to enter sampling dates with century information. The highest date that can be entered in **GRITS Database** and **GRITS SAGE** versions 1.0d or lower is December 31, 1999. A sample data set with dates that straddle the year 2000 boundary is entered in **GRITS SAGE** version 1.1 and successfully analyzed in both **GRITS Statistics** modules.

What is the Year 2000 Problem?
The seed of the "Year 2000 Problem" was planted in the early 1960s. Computer memory size was typically small and memory was expensive. One way to conserve memory is to store data in as compact a manner as possible. Consider ,for example, the date: "January 1, 1998". Each character in the date string "January 1, 1998" costs one byte. Therefore, storing the date with the month name spelled out is expensive and inefficient. Furthermore, since some month names are longer than others, some dates become more expensive to store than other dates (i.e., any date in January will require more storage than any date in May). Add to this the requirement of a month-name lookup table every time date calculations are required (i.e., January=1, February=2,...,December-12) and we have the worst possible storage scheme imaginable. A better scheme is to store the dates as MM/DD/CCYY (Month, Day, Century and Year). If we zero-pad the month (MM) and day (DD), all dates cost ten bytes. We can drop the cost to eight bytes if we omit the slash ("/") between the Month and Day and the slash between the Day and Year. A savings of two bytes may seem insignificant, but, consider a database record with five dates. For this single record this translates into a saving of ten bytes. For a small database with 1,000 records this is a saving of 10,000 bytes. The cost of date storage was reduced further when the century was dropped. Dates fields are now stored as YYMMDD and cost six bytes. The year and month were placed first in the date field for the purpose of sorting. Since memory, tape storage and hard drives were far more limited and much more expensive than they are today, these steps were considered appropriate. It literally saved computer owners the cost of extra memory, disk units and tape storage hardware. The savings is estimated at $30 billion since the 1960s[1]. The problem that this storage scheme causes is in the calculation of days that elapse between two given dates. When this calculation is made on two dates that have no century information, it is automatically assumed that the two dates are in the same century. Thus,

[1]Hall, David C. Lt. Col. (USAFR). *Year 2000 Problem: Infrastructure Aspects*. Wright-Patterson AFB, Dayton OH.

(99/12/31-99/12/30) equals one day ,but, (00/01/01-99/12/31) equals -100 years. If the result of such a calculation is used to compute simple interest on a loan the problem becomes apparent. **GRITS/STAT 5.0** uses a similar calculation to determine the horizontal axis on the time series plots.

The Year 2000 Problem and GRITS/STAT 5.0

In order to test **GRITS/STAT 5.0** with the Year 2000 problem the Copper observations in Table 1 were used. The first three sampling events occur in 1999 and the last three sampling events occur in the year 2000.

Sampling Date	MW-1 Background (ppb)	MW-2 Compliance (ppb)	MW-3 Compliance (ppb)
October 15, 1999	4.2	5.2	9.4
November 15, 1999	5.8	6.4	10.9
December 15, 1999	11.3	11.2	14.5
January 15, 2000	7.0	11.5	16.1
February 15, 2000	7.3	10.1	21.5
March 15, 2000	8.2	9.7	17.6

Table 1. Monthly Copper observations that straddle the century boundary.

Entering the data set in GRITS/STAT 5.0

GRITS SAGE Version 1.1 may be used to enter **GRITS/STAT 5.0** data with sampling dates that include the century. Note that **GRITS Database** and versions of **GRITS SAGE** prior to Version 1.1 do not prompt for the century when entering dates. The century will *always* default to "19" in **GRITS Database** and versions of **GRITS SAGE** prior to version 1.1.

1. Start **GRITS/STAT 5.0** and select **GRITS SAGE**.

2. Select the System Utilities and Maintenance... option was selected from the Data System Main Menu.

3. Select the Create Empty GRITS Databases from the System Utilities and Maintenance menu.

4. At the Directory Name prompt type **C:\GRITS500\YR2000 <Enter>**. This will create the C:\GRITS500\YR2000 directory on your hard drive and populate it with empty **GRITS/STAT 5.0** database files. Press **<Esc>** to return to the Data System Main Menu.

5. Select the Parameter, Date x Well Spreadsheet from the Data System Main Menu. A blank spreadsheet should appear on your screen. Press <**Insert**>.

6. Select the Facility option from the Add pop-up menu. The New Facility dialog will appear on your screen. Make the entries shown in Figure 1 in the New Facility dialog and press <**Page Down**>.

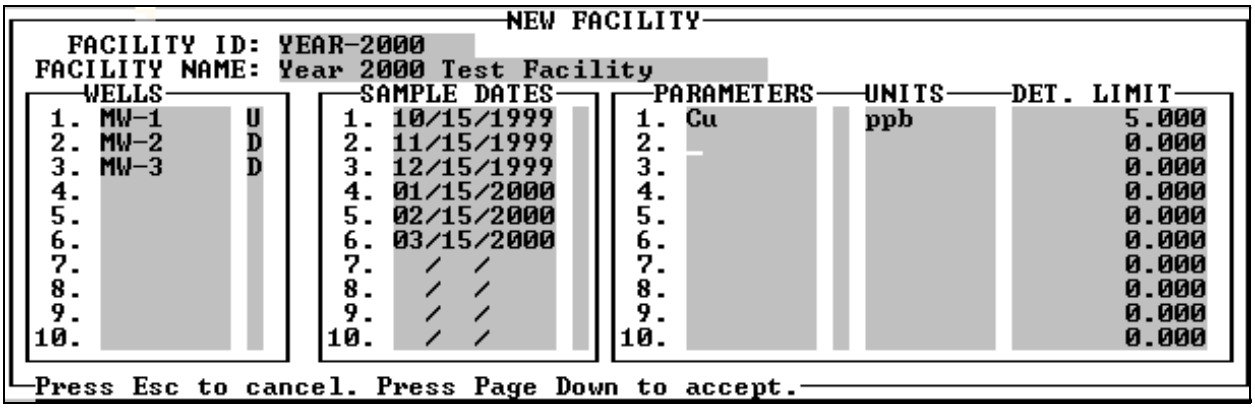


Figure 1. The New Facility dialog for the data in Table 1.

7. Enter the Copper observations from Table 1. When you have completed the data entry your spreadsheet should look like Figure 2.

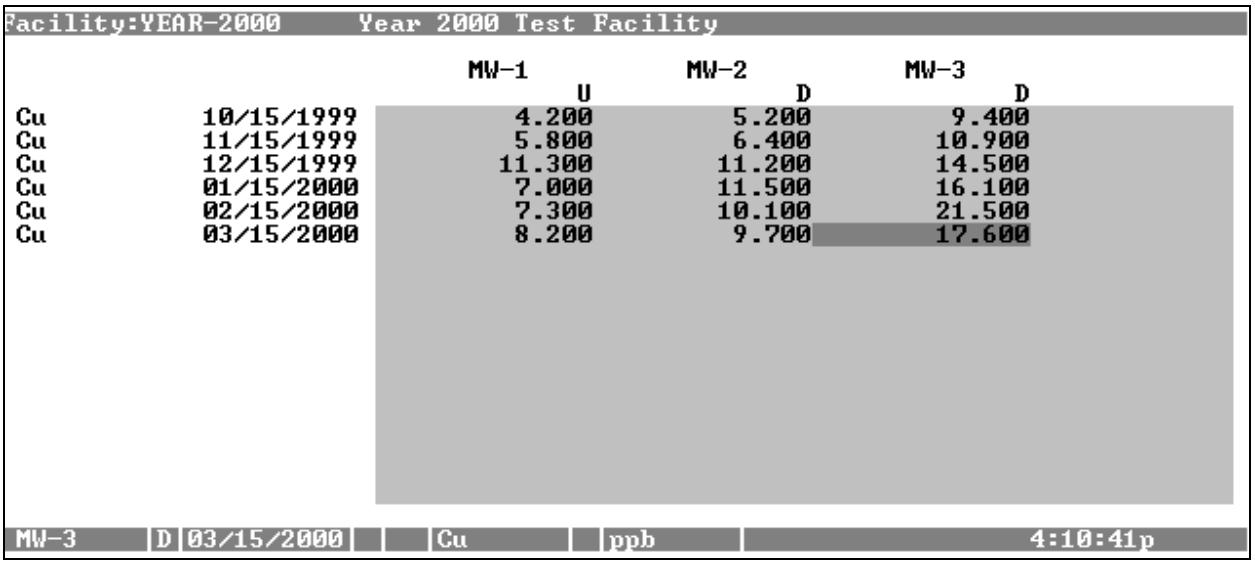


Figure 2. The spreadsheet screen for the data in Table 1.

8. Press <**Esc**> to back out of the spreadsheet and exit **GRITS SAGE**.

9. Launch the **GRITS Database** module. Select the Select a Facility option from the DATA MENU. Press **<F2>**. At the Source Path prompt type: **<C:\GRITS500\YR2000> <Enter>**

   The data directory in **GRITS/STAT 5.0** is now set to C:\GRITS500\YR2000. Note that selecting or creating a directory in **GRITS SAGE** *does not* set the data directory for the other **GRITS/STAT 5.0** modules.

10. Exit **GRITS Database**. Launch one of the GRITS Statistics modules and load the data for analysis as you would normally (for detailed instruction on loading and creating data sets in the **GRITS Statistics** modules see section 5.3.1 of the GRITS/STAT 5.0 Manual).

Analysis in GRITS Statistics Intervals Module

A one-tailed upper 95% Parametric Prediction Interval was constructed on the Copper observations in background well MW-1. The compliance well observations were compared to the upper limit of the interval. The analysis was performed using the April 22, 1997 version of the **GRITS Statistics Intervals Module**.

Figures 3 and 4 show a time-series and box plot of the Year 2000 Copper data set as produced by GRITS Statistics.

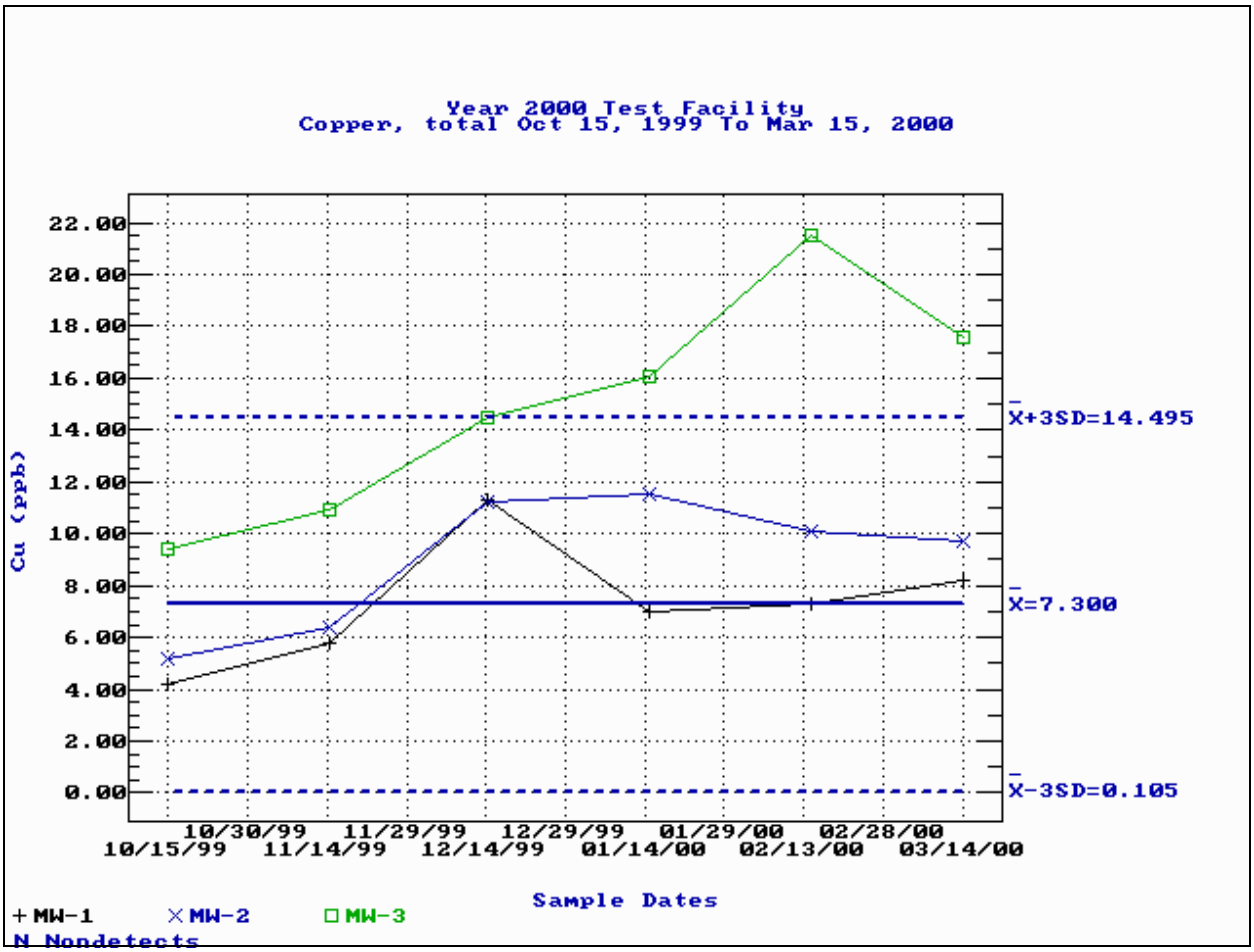


Figure 3. Time-series plot of the Year 2000 data set as produced by GRITS Statistics Intervals Module.

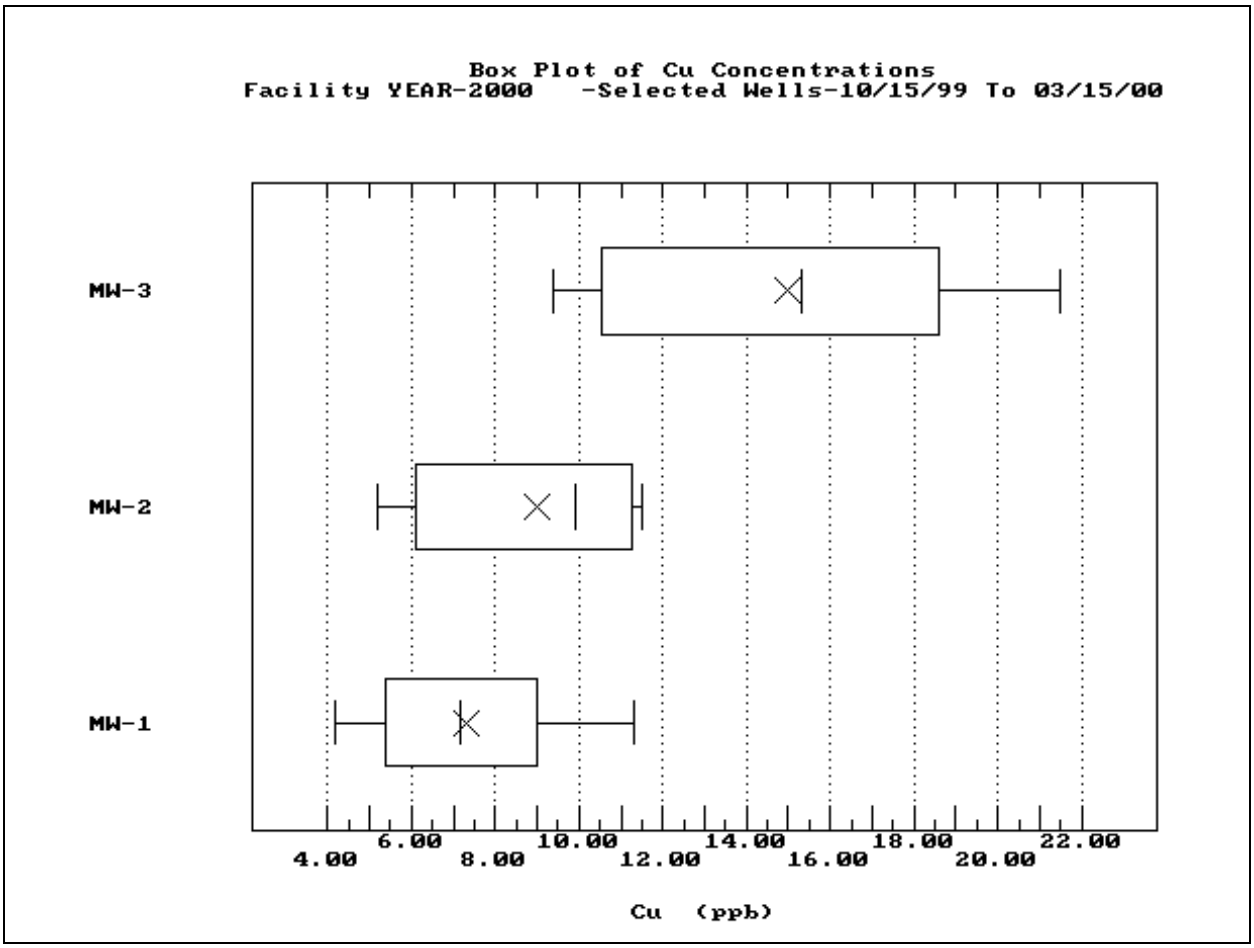
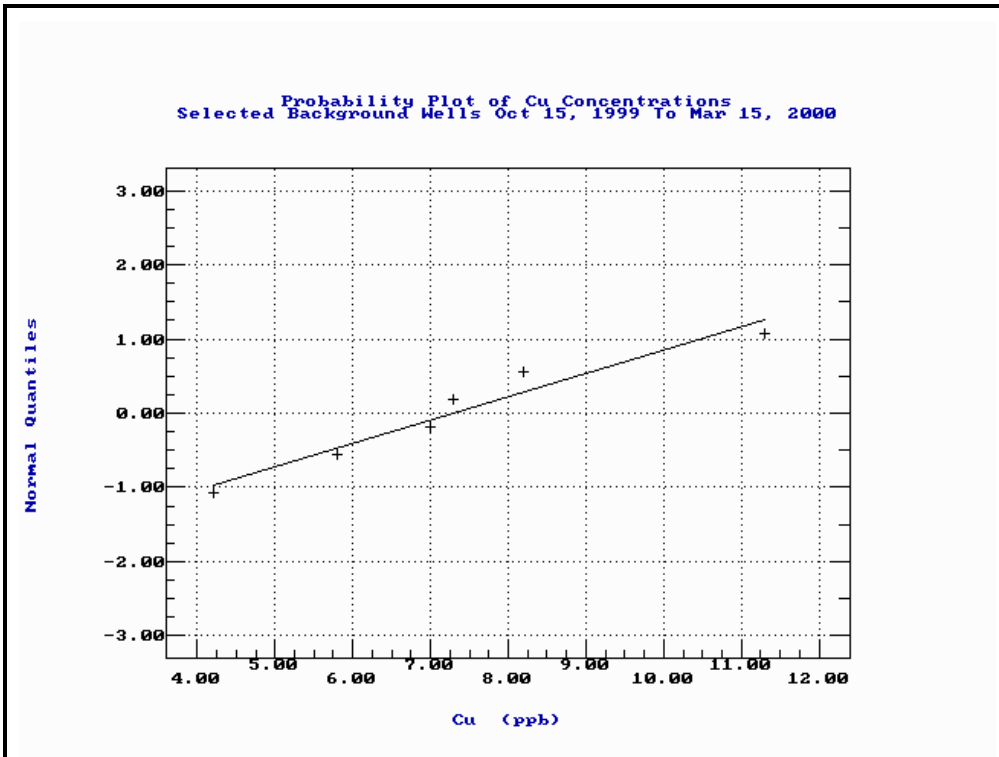


Figure 4. Box Plot of the Copper observations in the Year 2000 data set. Note that since dates are not used for either axis, the Year 2000 problem does not affect Box Plots produced by **GRITS/STAT 5.0**.

Since we are constructing a Parametric Prediction Interval on the observation in well MW-1, we may want to test these observations for Normality. A Probability Plot of the observations in well MW-1 is shown in Figure 5. Since the Probability Plot Correlation Coefficient (r=0.973) exceeds the 5% critical value for sample size 6 $R_{0.5,\,6}$=0.890), the sample shows no significant evidence of non-Normality by the Probability Plot Correlation Coefficient test and we may proceed to construct the interval. The results of the Parametric Prediction Interval test are shown in Figures 6 and 7. Since the last four observations in compliance well MW-3 exceed the upper prediction limit of 12.52 ppb there is statistically significant evidence that these observations exceed the background well observation in well MW-1. These results are graphically illustrated in Figure 7.



Observations: 6
Mean: 7.3000
Std Dev: 2.3983
Minimum: 4.2000
Maximum: 11.3000
Median: 7.1500
Skewness: 0.5008
r: 0.9733
$R_{(\alpha=0.01,N)}$: 0.835
$R_{(\alpha=0.05,N)}$: 0.890

Figure 5. Probability Plot on the background observations in well MW-1. Note that Probability Plots do not use dates on either axis and are ,therefore, not affected by the Year 2000 problem.

Parametric Prediction Interval
Report Printed December 15,1997

Page 1

Facility:Year 2000 Test Facility
Parameter:Copper, total(CAS Number:7440-50-8)

ONE-TAILED UPPER PARAMETRIC PREDICTION INTERVAL

Observations (n): 6
Shapiro-Wilk (W): 0.9641
Critical W,"=0.01: 0.7130
Mean: 7.300 ppb
Std Dev: 2.398 ppb
DF: 5
Conf. Level (1-"): 0.9500
Future Samples (k): 1
t+ 1 - " ,: 2.0150
*) *
. k -
Kappa: 2.1765

UL: 12.520 ppb
LL: -∞

BACKGROUND TO COMPLIANCE WELL COMPARISON

Well:MW-2

Sample Date	Observation
10/15/99	5.200 ppb
11/15/99	6.400 ppb
12/15/99	11.200 ppb
01/15/00	11.500 ppb
02/15/00	10.100 ppb

```
-----------------------------------------
     03/15/00                   9.700 ppb
-----------------------------------------

 Well:MW-3

    Sample Date                Observation
     10/15/99                   9.400 ppb
-----------------------------------------
     11/15/99                  10.900 ppb
-----------------------------------------
     12/15/99                  14.500 ppb *
-----------------------------------------
     01/15/00                  16.100 ppb *
-----------------------------------------
     02/15/00                  21.500 ppb *
-----------------------------------------
     03/15/00                  17.600 ppb *
-----------------------------------------
Report Produced by GRITS/STAT 5.01
```

Figure 6. Textual results of the Parametric Prediction Interval method in GRITS/STAT 5.0 on the Year 2000 data set.

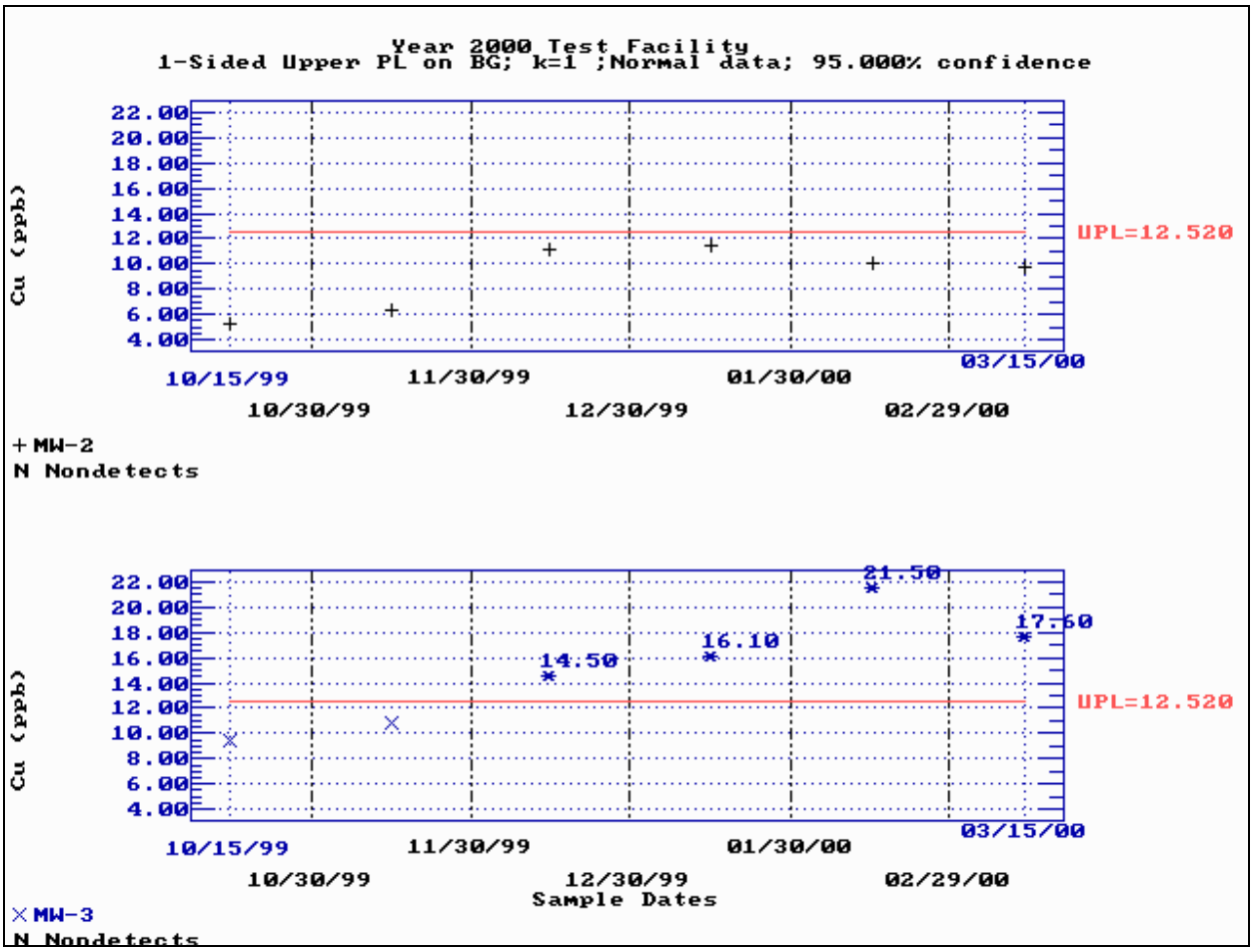


Figure 7. Graphical display of the Parametric Prediction Interval results produced by **GRITS/STAT 5.0**.

The results of a one-sided upper 95% Parametric Tolerance Interval with 95% coverage

constructed around the background observations in well MW-1 are shown below in Figures 8 and 9.

```
                        Parametric Tolerance Interval
                        Report Printed December 15,1997

                                                                        Page 1
 Facility:Year 2000 Test Facility
Parameter:Copper, total(CAS Number:7440-50-8)

ONE-TAILED UPPER PARAMETRIC PREDICTION INTERVAL

    Observations (n):      6
    Shapiro-Wilk (W):  0.9641
   Critical W,"=0.01:  0.7130
                Mean: 7.300 ppb
             Std Dev: 2.398 ppb
                  DF:      5
   Conf. Level (1-"):  0.950
            Coverage:  0.950
               Kappa:   3.7080

                  UL: 16.193 ppb
                  LL: -∞


BACKGROUND TO COMPLIANCE WELL COMPARISON

 Well:MW-2

    Sample Date               Observation
     10/15/99                   5.200 ppb
     11/15/99                   6.400 ppb
     12/15/99                  11.200 ppb
     01/15/00                  11.500 ppb
     02/15/00                  10.100 ppb
     03/15/00                   9.700 ppb

 Well:MW-3

    Sample Date               Observation
     10/15/99                   9.400 ppb
     11/15/99                  10.900 ppb
     12/15/99                  14.500 ppb
     01/15/00                  16.100 ppb
     02/15/00                  21.500 ppb *
     03/15/00                  17.600 ppb *
```

Figure 8. Textual results of the Parametric Tolerance Interval method produced by **GRITS/STAT 5.0** for the Year 2000 data set.

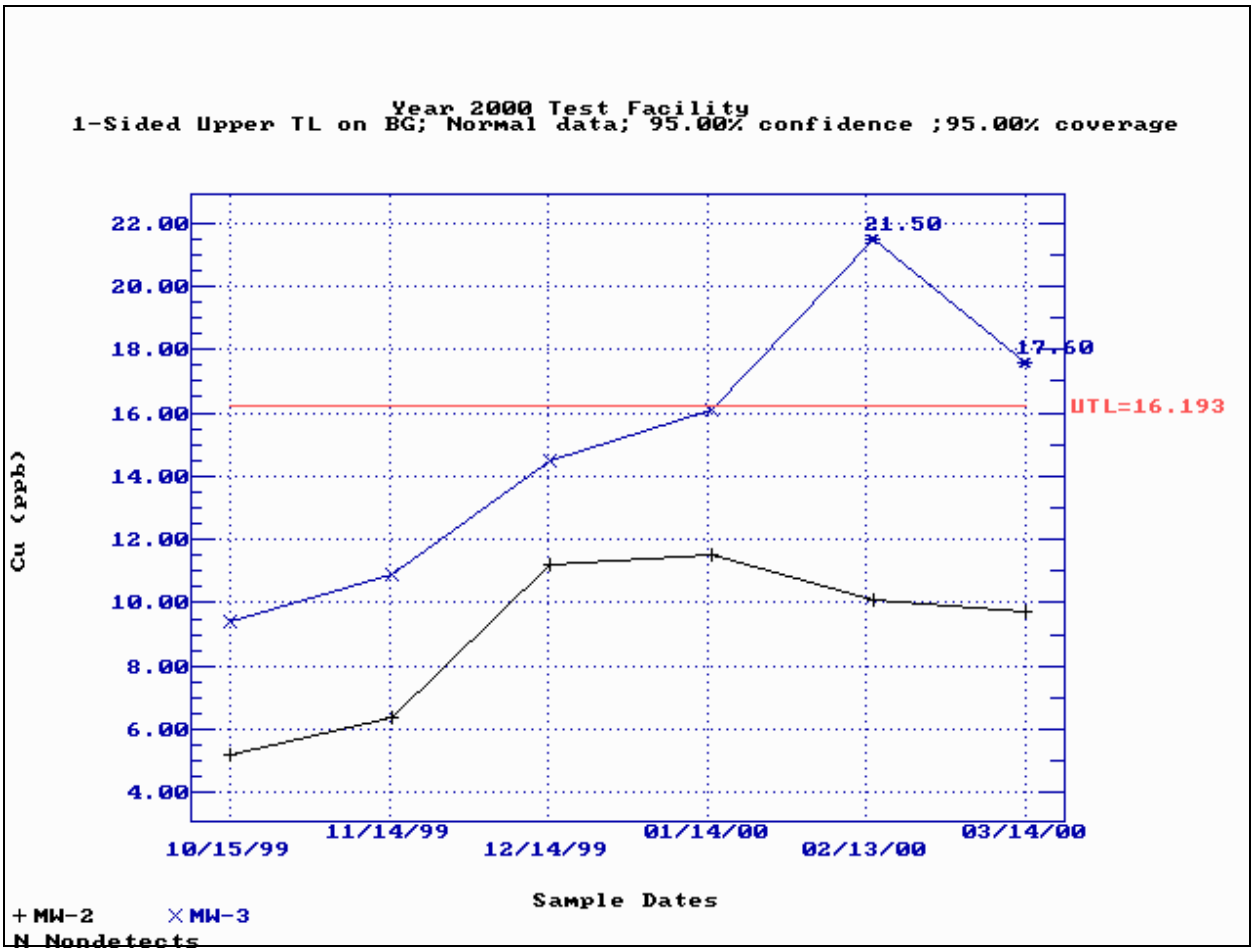


Figure 9. Graphical results of a one-sided upper 95% Parametric Tolerance Interval with 95% coverage as produced by **GRITS/STAT 5.0** on the Year 2000 data set.

Since the last two samples in well MW-3 exceed the upper-tolerance limit of 16.19 ppb, there is significant evidence that the concentration distributions of well MW-1 (background) and well MW-3 (compliance) are different and conclude there is significant evidence of contamination in well MW-3.

Analysis in the GRITS Statistics ANOVA, Two-Sample, Control Chart Module

A Shewhart-CUSUM Control Chart was produced for the Year 2000 data set of Table 1 using the April 18, 1997 version of the **GRITS Statistics ANOVA, Two-Sample, Control Chart Module** in the current release of **GRITS/STAT 5.0**. The observations in the background well MW-1 were used to compute a baseline mean and standard deviation for a control chart on well MW-3. The results of the analysis are shown in Figures 10 and 11.

```
+Data Mode:Original-------------------------------------------------------------+
|    Facility:YEAR-2000    Year 2000 Test Facility                             |
| Permit Type:Detection                                                        |
| Constituent:Cu       Copper, total                                           |
|    Starting:Oct 15, 1999          Ending:Mar 15, 2000                        |
+------------------------------------------------------------------------------+
                      Mean (μ):       7.300  ppb
             Std Deviation (F):       2.398  ppb
  Decision Internal Value (h):        5.00
          Reference Value (k):        1.00
  Shewart Control Limit (SCL):        4.50

* Indicates out-of-control process via (Zi>SCL) or (Si>h)
Date        T       N     Mean Xavg    Std Mean Z          Z-k      CUSUM S
 Oct 15     1       1          9.40          0.88         -0.12         0.00
 Nov 15     2       1         10.90          1.50          0.50         0.50
 Dec 15     3       1         14.50          3.00          2.00         2.50
*Jan 15     4       1         16.10          3.67          2.67         5.17
*Feb 15     5       1         21.50          5.92          4.92        10.09
*Mar 15     6       1         17.60          4.29          3.29        13.39
```

Figure 10. Textual results of the Shewhart-CUSUM control chart method as reported by **GRITS/STAT 5.0** for the Year 2000 data set.

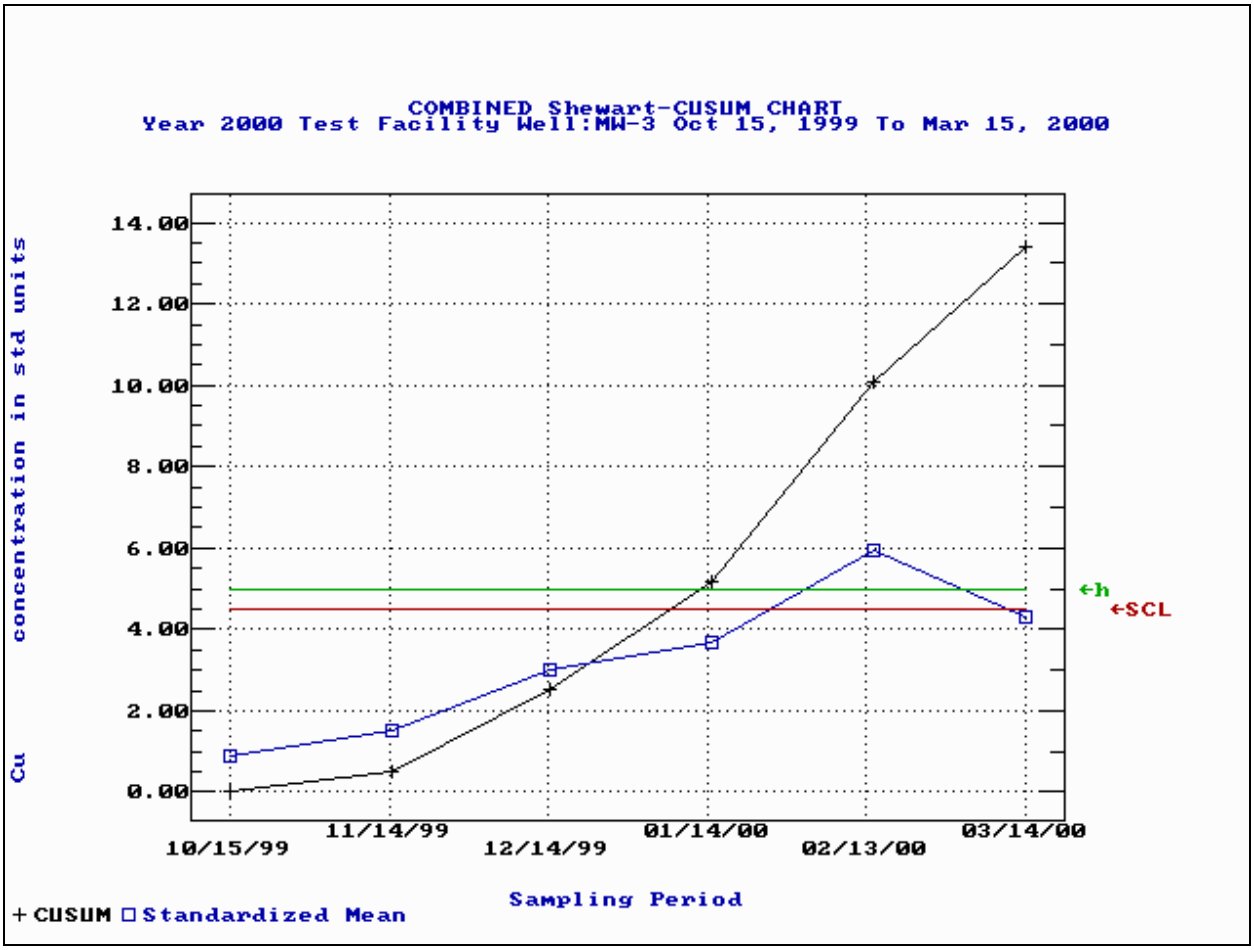


Figure 11. Shewhart-CUSUM Control Chart produced by **GRITS/STAT 5.0** for the Year 2000 data set.

Since S_i exceeds the decision internal value (h=5.0) for all three of the sample collected in the year 2000 at well MW-3, we conclude that the Shewhart-CUSUM Control Chart for well MW-3 is

"out-of-control" and there is statistically significant evidence of contamination at well MW-3.

Conclusion

The presence of a data set that crosses the 1999-2000 year boundary presents no problem to the statistical modules in **GRITS/STAT 5.0**. It may be desirable to include the century in the display of dates in reports and plots produced by future versions of the statistics modules. The chief difficulty lies in data entry. Users who need to enter dates outside of the 20th century must use **GRITS SAGE** version 1.1 or higher. **GRITS SAGE** version 1.1 was used to enter the data set discussed in this document and will be released in December 1997.

Initial tests of the **GRITS Advanced Reporter** and **GRITS Reporter** modules, however, were not encouraging. Both modules seem to have a problem reporting results with sample dates outside of the 20th century. The **GRITS Database** module will not currently load observations with a date outside of the 20th century. The reports in the **GRITS SAGE** version 1.1 report menu do, however, work properly for dates outside of the 20th century. The Year-2000 compliance for each of the modules in **GRITS/STAT 5.0** is summarized in Table 2.

GRITS/STAT 5.0 Module	**Year 2000 Compliant?**
GRITS Database	No
GRITS Reporter	No
GRITS Statistics,Intervals Module	Yes
GRITS Statistics,ANOVA,Two-Sample,Control Chart	Yes
GRITS Exporter	No
GRITS Utilities	No
GRITS Advanced Reports	No
GRITS Sage	Version >= 1.1 Yes Version <= 1.0d No

Table 2. Year 2000 compliance for the **GRITS/STAT 5.0** Modules.

Due to the data driven nature of the **GRITS SAGE** (i.e., screens are not hard-coded ,but, saved in a database), the changes required to make **GRITS SAGE** Year 2000 compliant were inexpensive (≈ 1 hour). Changes to the other **GRITS/STAT 5.0** modules, however, could be quite expensive.